Journal of Information Technology and Cyber Security 1(1) January 2023: 22-30 EISSN 2987-386X | PISSN 2987-3878
DOI: 10.30996/jitcs.7676

The Utilization of Information System for Crime Rate Modelling in Surabaya Using K-means

Supangat Supangat [1,*] and M. Mudhafiq Sholiq [2]

[1] Department of Information Systems and Technology, Universitas 17 Agustus 1945 Surabaya, Indonesia
[2] Department of Informatics Engineering, Universitas 17 Agustus 1945 Surabaya, Indonesia
* Corresponding author: supangat@untag-sby.ac.id

Received: 12 December 2022
Accepted: 09 January 2023
Revised: 04 January 2023
Available online: 07 February 2023

To cite this article: Supangat, S., & Sholiq, M. M. (2023). The Utilization of Information System for Crime Rate Modelling in Surabaya Using K-means. *Journal of Information Technology and Cyber Security*, 1(1), 22–30. https://doi.org/10.30996/jitcs.7676

Abstract
This study aims to model the crime rate in the city of Surabaya using the k-means clustering method. The data used is crime data that occurred in Surabaya in previous years, which includes the type of crime, location of crime, and crime rate. The k-means clustering method is used to classify crime data in the Surabaya area for 2020-2022 consisting of cluster 3, namely areas with moderate crime rates covering 6 sub-districts (1,260 cases), cluster 1 with areas with high crime rates, namely 12 sub-districts with 2,363 cases, and cluster 2 areas with low crime rates consisting of 13 districts with 2,178 cases based on data on the number of crimes. The geospatial visualization system is used to visually display modeling results, making it easier for interested parties to identify the location of a crime. The results of this study are expected to provide useful information for interested parties, such as the police and the community, in taking preventive action regarding crime rates in Surabaya.
Keywords: clustering method, crime modeling, geospatial data visualization, k-means, unsupervised learning.

1. Introduction
Surabaya is the second largest city in Indonesia after the capital city of Jakarta. Surabaya is a regional city with a large population from various regions. This causes an imbalance between the number of jobs and the population which eventually causes some residents who are desperate and commit crimes to meet their needs. Data shows that there were 58 cases of street crime in just two months, January-February in 2022. The Chief of Surabaya Big City Resort Police (Polrestabes) stated that the street crime cases included 34 cases of motorcycle theft and 24 cases of theft with violence or robbery (Arfani & Nashrullah, 2022).

Crime rate in Surabaya has increased from time to time, especially motorcycle theft (Arsista, 2022; Himawan, 2023), so that extra action and handling is very important to do as an effort to overcome these problems (Anshori & Misbachudin, 2017). Street crime incidents that occur can be caused by the absence of information about areas with high crime rates (Astuti, 2018), so that the community and the police have not been able to take effective preventive measures. The same problem related to criminality has been widely studied, among others: 1) Bindosano et al. (2022) proposed a search for crime-prone areas in Jayapura City, Papua Province, Indonesia to reduce crime with Crime Through Environment Design (CPTED) and looking for a relationship between the perception of security of existing citizens and CPTED variables; 2) Nurman (2007) proposed a web-based crime profile mapping information system that can display conventional crime information in Bogor City, West Java Province, Indonesia. The information displayed is in the form of text, map, and graphical data (Hapsari & Widodo, 2017); 3) Rahayu et al. (2014) proposed a clustering technique to determine the potential for regional crime in Banjarbaru City, South Kalimantan Province, Indonesia based on alignment (Hapsari & Widodo, 2017); 4) Gunawan & Aditya (2019) proposed the use of geovisual analytics of crime using social media data to identify patterns and movements of crime incidents in Jakarta, Indonesia; 5) Setiawan et al. (2019) proposed a geographical approach to analyze the relationship between crime and accessibility in Sumur Bandung as the area with

the highest crime rate in Bandung City, West Java Province, Indonesia; 6) Nurjoko et al. (2020) developed a geographic information system for mapping areas with high crime rates using clustering techniques; 7) Mulyani et al. (2020) developed web-based applications that combine the k-means algorithm to group vulnerable areas and Geographic Information Systems (GIS) to map crime-prone areas. Five parameters are used in the application developed by Mulyani et al. (2020), namely: theft, molestation, rape, women and child protection cases, and fraud.

Based on the previous studies, we proposed a system based on the k-means algorithm combined with geospatial visualization. The system can later provide information to the public and the police about which areas have the potential for crime, so as to increase the level of alertness, anticipation, and eventually be able to help reduce the risk of crime.

2. Methods

2.1. Data collection

Table 1
Data of the number of crimes in Surabaya.

District	Number of Crime Cases (Per Year)		
	2020	2021	2022
Asemrowo	85	68	39
Benowo	58	72	39
Lakarsantri	75	47	50
Pakal	60	66	68
Sambikerep	56	91	49
Suko Manunggal	86	72	44
Tandes	64	66	48
Dukuh pakis	57	81	51
Gayungan	70	56	52
Jambangan	72	82	38
Karang Pilang	72	68	38
Sawahan	77	57	35
Wiyung	77	78	28
Wonocolo	65	52	40
Wonokromo	90	77	55
Gubeng	59	43	47
Gunung Anyar	58	62	32
Mulyorejo	71	65	53
Rungkut	96	92	50
Sukolilo	78	60	28
Tambaksari	82	85	47
Tenggilis Mejoyo	75	55	32
Bulak	62	62	50
Kenjeran	66	57	41
Krembangan	72	66	59
Pabean Cantian	73	83	57
Semampir	72	57	50
Bubutan	76	72	56
Genteng	88	56	48
Simokerto	74	97	52
Tegalsari	98	62	54

In this study, data collection was carried out based on interviews, data collection from public and private data, and literature studies. Data collection through interviews was carried out with the police in Surabaya City regarding the number of crimes, locations where crimes often occur, and types of crime cases from 2020 to 2022. Before the interview, it is necessary to prepare data by giving several questions related to research problems to the resource person, namely the Chief of Surabaya Polrestabes. Public data is obtained from the official website of the government and research institutions related to the problems in this study. Private data is obtained from the police information system and population information system, literature studies related to previous research relevant to this research. The results of interviews with resource person are data on the number of cases, types of crimes, and areas where cases occurred as shown in Table 1.

2.2. Data preprocessing

The stage after data collection is data reduction and data presentation. The data reduction stage is carried out by reducing the amount of data not needed in this study so that only important data remains. Data reduction in this study is carried out by dividing data into certain categories and themes (Rijali, 2018).

2.3. K-means

The k-means algorithm is one of the unsupervised learning machine learning algorithms that can divide data into groups (clusters) that have similarities or something in common. The first step in k-means is to determine the number of clusters (k), e.g., if one wishes to group as many as three groups, then $k = 3$. The next step is to determine the centroid (center of cluster) for each cluster. Usually, the centroid is randomly selected. After the centroid selection, calculate the closest distance from each data object with centroids, in this study Euclidean distance is used to calculate the distance using Eq. (1) (Larose & Larose, 2014),

$$d(x, y) = \sqrt{\sum_i (x_i - y_i)^2} \quad (1)$$

where $d(x, y)$ is Euclidean distance between x and y, while $x = x_1, x_2, \dots, x_m$, and $y = y_1, y_2, \dots, y_m$ represents the m attribute value or data points from two data, in this case, the object data and the centroid in which its distance will be calculated.

Eq. (1) with $k = 3$ returns dc_1 which is the distance to each data object with Cluster 1, dc_2 is the distance to each data object in Cluster *2,* and dc_3 is the distance to each data object in Cluster 3. After the distance dc_1, dc_2, and dc_3 are calculated, one can determine the closest distance from the object to centroids. Once the closest distance is determined, the object is assigned to the cluster with the closest distance. This process is carried out repeatedly until it converges by updating or recalculating the latest centroid value using Eq. (2)(Prasetyo, 2014),

$$c = \frac{1}{n_i} \sum x_i \quad (2)$$

where c is the updated centroid. n_i is the number of data points in the cluster i. x_i is the feature vector of each data point in the cluster. $\sum x_i$ is the sum of all feature vectors in the cluster i.

2.4. Software development

In this research, software development uses the Rapid Application Development (RAD) method because it can be used easily, save time, which is within 30-90 days, allows cost savings, and good quality software results. According to Kendall and Kendall (2011), RAD consists of three main phases that involve users and analysts in the assessment, design, and implementation process.

1. Requirements Planning Phase: In the requirements planning phase, users and analysts meet to identify system objectives as well as information requirements arising from those goals.
2. RAD Design Workshop: In this phase, users respond to actual working prototypes and analysts refine designed modules based on user responses.
3. Implementation Phase: In this phase, analysts work intensively with users to design business or nontechnical aspects of the system. Once these aspects are agreed upon and the system is built and refined, a new system or part of the system is tested and introduced into the organization.

2.5. System flowchart

Stakeholders involved in utilizing this system consist of 1) Admins who act as system managers and geospatial data managers; 2) Polrestabes which acts as the source of crime data in Surabaya; 3) Users who act as the main role in operating the spread of crime zoning information system; and 4) The system itself, which acts as a medium of use, information, management and calculation of zoning data for regional spread.

Table 2
Functional Requirements.

Code	Functional Requirements (FRs)
FR-001	Access the main page of the website.
FR-002	Admin logs in on the admin page.
FR-003	Admin modifies user account (add, change, delete).
FR-004	Admin modifies the spread of crime or crime distribution data (add, change, delete).
FR-005	Admin performs crime data clustering with k-means clustering.
FR-006	Admin can view reports entered by users.
FR-007	User can see the location of the spread of crime displayed by the system in the form of a map with k-means.
FR-008	User can see a list of crime distribution data in Surabaya area
FR-009	User can see detailed data on the spread of crime in the Surabaya area.
FR-010	User can search crime data with crime type keywords.
FR-011	User can report crime events on the report menu.
FR-012	The system can redirect from geospatial websites to Google Maps.

The system workflow process is designed with architecture that can support the functionality of the system. The workflow process of the system is illustrated in Fig. 1. The workflow of the system starts with the main web page that displays a map of the spread of crime in Surabaya featuring data such as the

number of crime cases, the number of victims, and the types of cases that have occurred. The spread of crime zoning is classified into three regional clusters, namely areas with high crime rates, areas with medium crime rates, and areas with low crime rates.

2.6. System requirements analysis

At this stage, we conduct an analysis of software requirements, namely functional requirements and non-functional requirements. Functional requirements are requirements that must be met so that a system can run according to the desired purpose. Non-functional requirements consist of service or function limitations on the system, development process limitations, and standardization of system and user requirements. Functional requirements are presented in Table 2 and non-functional requirements are presented in Table 3.

2.7. Actors' scenarios

Scenarios are used to describe the actors or stakeholders running the system. The first stakeholder is the admin or Polrestabes. The admin accesses the main page, logs in, enters the main admin page, modifies the crime distribution area data, then determines the k-means cluster, modifies crime case data consisting of number data, victim data, type data, viewing user reports, and logging out. The second stakeholder is the user. The user accesses the main page, searches for crime data, views the zoning of crime distribution areas in the form of maps, views detailed crime case data, accesses the report menu, and fills out the crime report data.

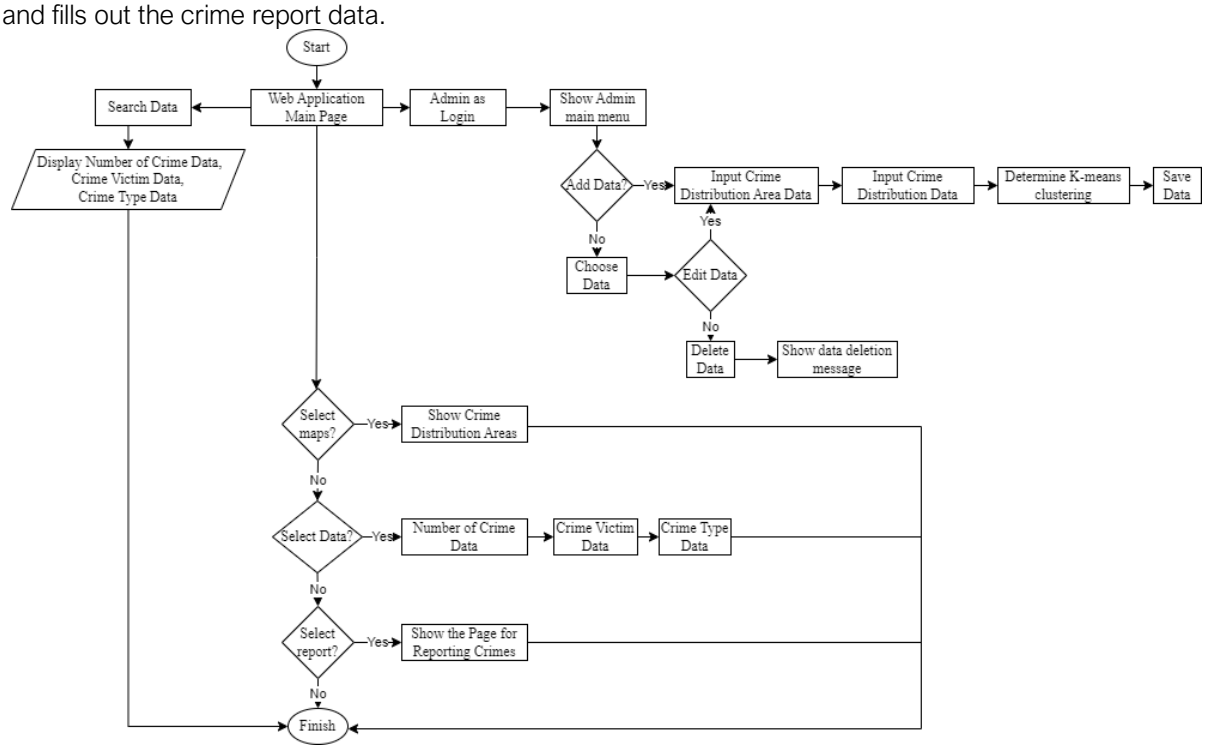

Fig. 1. System flowchart.

Table 3.
Non-functional requirements.

Code	Non-functional Requirements (NFRs)
NFR-001	The system can run normally with minimal errors.
NFR-002	The system can be accessed either by admins, Polrestabes, or other visitors anywhere as long as they are connected to the internet.
NFR-003	The system can operate for 24 hours nonstop.
NFR-004	The system has a smooth and responsive interface.
NFR-005	The system can be run by several web browser software, including Internet Explorer, Google Chrome, and Mozilla Firefox.
NFR-006	The system has guaranteed data security.

2.8. Database

In the initial design, the database is conceptual in the form of an Entity Relationship Diagram (ERD) that connects entities with relationships. The required entities are admin, report, and region. The admin entity is the main part that will function in the system operator. Admins can manage multiple crime spread areas. A report can be accessed by multiple admins and an admin can access multiple reports.

2.9. User Interface

The graphical user interface is designed to display every data needed by the user by accessing the data with an easy-to-understand or user-friendly presentation. The design of the graphical interface that displays the main page when users and admins access the website is depicted in Fig. 2.

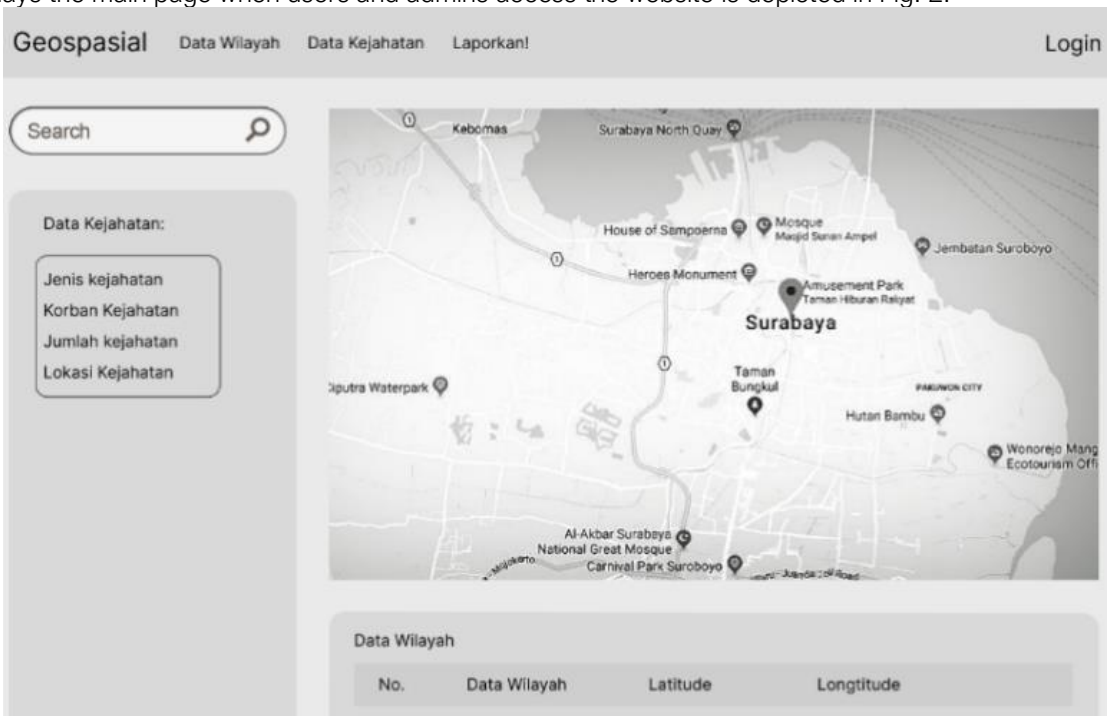

Fig. 2. Main page of the proposed system.

2.10. Implementation of the proposed system

The next stage is the implementation process. The system has a user's main page, Admin dashboard, data management page in terms of adding, changing, removing, and displaying crime distribution data, crime grouping page to present maps and groupings based on k-means. One illustration of the system developed at the implementation stage is the crime grouping page which is presented by employing a map as in Fig. 3.

3. Results and Discussion

The testing and evaluation conducted in this study aim to find out whether the proposed system is ready to be deployed to the users. Testing in this study uses the smoke testing method, which includes running several predetermined scenarios and ensuring that the system or application does not experience fatal failures during the testing process. The test scenario in this study is presented in Table 4. The k-means results of each region are presented in Table 5 and Table 6.

Based on Table 5, it is known that the crime distribution cluster in the Surabaya area in 2020-2022 consists of three clusters. Cluster 1, with areas with high crime rates, consists of 12 sub-districts with 2,363 cases. Cluster 2, with areas with low crime rates, consists of 13 sub-districts with 2,178 cases. Cluster 3, with areas with medium crime rates, consists of 6 sub-districts with 1,260 cases.

4. Conclusions

This study proposes a system that combines geospatial technology as information on crime points in Surabaya City with the k-means algorithm. The purpose of the research is to determine the level of vulnerability of an area in Surabaya City. The system is designed to be accessible to the public so that they know the map of the spread of crime and make it easier to report incidents/behavior of criminal acts in the areas of Surabaya, those that have occurred and are happening, by accessing this application page online without having to come to the Polrestabes office or the Sector Police (Polsek); From the results of the cluster grouping, it is obtained that the crime rate in Surabaya City in 2020-2022 is divided into three regional groupings, namely Cluster 1 area with high crime rate, Cluster 2 area with low crime rate, and Cluster 3 area with medium crime rate. This research has limitations, so further research is still needed, namely in terms of area size, handling weaknesses of k-means, and measuring performance evaluation of k-means.

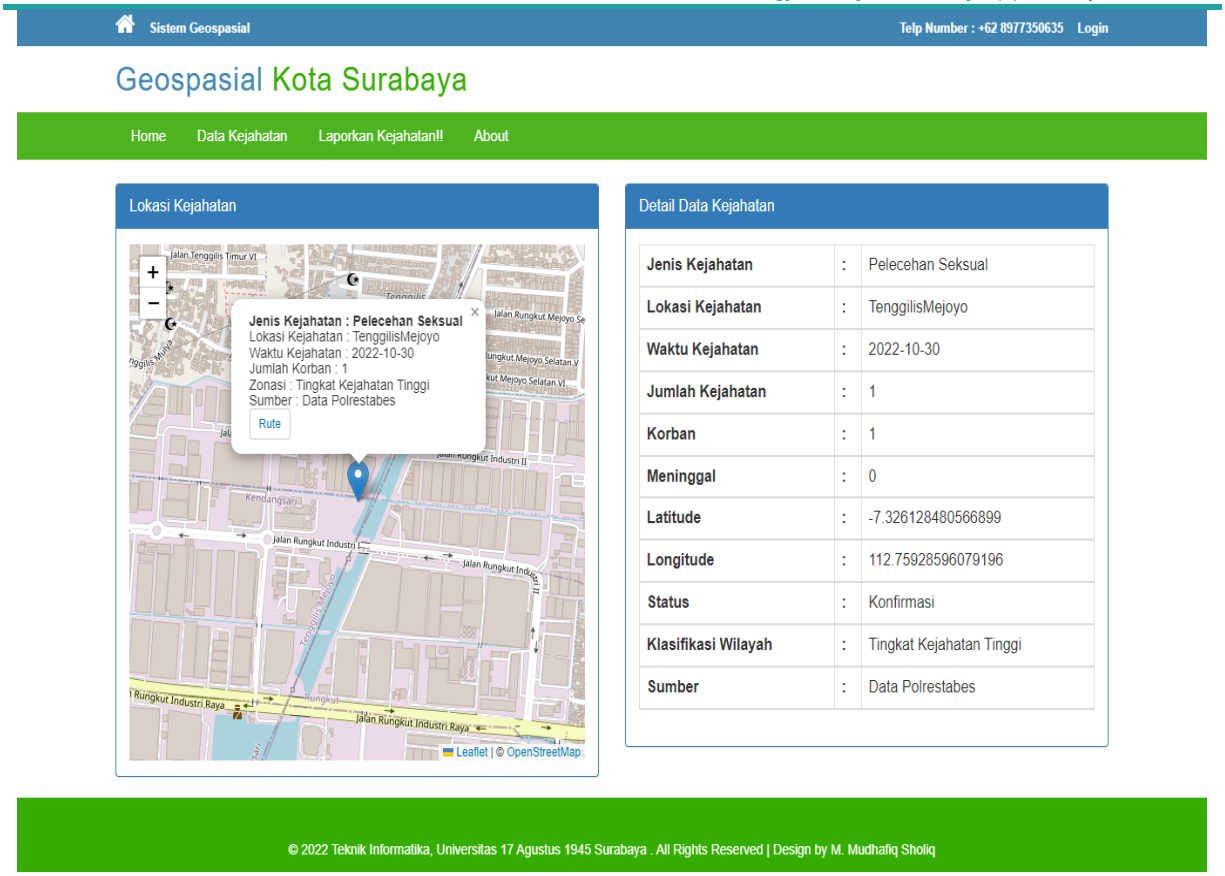

Fig. 3. Crime clustering on the map.

Table 4
Smoke testing scenarios.

#	Test Scenario	Test Case	Expected Result	Test result status
1.	Access the main page of the website.	Type in the browser's Address Bar "localhost/geospasial-surabaya".	The system can display the main page of the website.	Valid
2.	Admin logs in on the admin page.	Fill out the username and password form and click "log in".	The system accepts logging in to the home admin.	Valid
3.	Admin modifies user account (add, change, delete).			
	Admin adds user data.	Fill out the user data input form with the name "untag" and click "save".	The system successfully saves and displays the user account	Valid
	Admin changes user data.	Change one of the users "untag" then click save.	The system successfully updates the user account.	Valid
	Admin deletes user data.	Delete one of the user accounts "untag" then click the action "delete".	The system successfully deletes the selected user data.	Valid
4.	Admin modifies the spread of crime or crime distribution data (add, change, delete).			
	Admin adds distribution data.	Fill out the entire distribution data input form and click "save".	The system successfully saves the data and displays the entered scatter data.	Valid
	Admin changes distribution data.	Change one of the distribution data "data" and click "save"	The system successfully updates the distribution data.	Valid
	Admin deletes distribution data.	Delete one of the distribution data "test data" and click the "delete" action.	The system successfully deletes the selected distribution data.	Valid
5.	Admin performs crime data clustering with k-means clustering.			
	Admin uploads dataset on k-means menu	Select Excel file then upload it to dataset form	The system successfully displays the dataset uploaded by the admin	Valid
	The admin determines the number of clusters and centroids.	Enter the number of clusters and select the centroid then click "save".	The system displays forms for centroid and cluster determination.	Valid

(*continued on next page*)

Table 4. (*continued*)

#	Test Scenario	Test Case	Expected Result	Test result status
	Admin accesses k-means process menu	Access the k-means process menu and view looping data	The system successfully displays the k-means looping process	Valid
	Admins accesses the cluster results menu	Access the clustering menu and view the results of the k-means data cluster	The system successfully displays the results of the k-means data clustering cluster	Valid
6.	Admin can view reports entered by users.	Access the reports menu and view report data.	The system successfully displays the report menu page on the admin dashboard.	Valid
7.	User can see the location of the spread of crime displayed by the system in the form of a map with k-means.	Access the main menu of the website and view the mapping in the form of maps.	The system successfully displays mapping crime distribution data in the form of geographical maps.	Valid
8.	User can see a list of crime distribution data in Surabaya area.	Access the crime data menu on the main display.	The system successfully displays data on the number of crime cases.	Valid
9.	User can see detailed data on the spread of crime in the Surabaya area.	Select one of the crime records and click the "detail" action.	The system successfully displays the details of the spread of crime and displays the location point in the form of a map.	Valid
10.	User can search crime data with crime type keywords.	Input the keyword "drugs" in the search form.	The system successfully displays data search results based on keywords entered by the user.	Valid
11.	User can report crime events on the report menu.	Fill out the entire crime report data form, then click "save"	The system displays a "Data Saved Successfully" notification	Valid
12.	The system can redirect from geospatial websites to Google Maps.	Select one of the crime records, then click "route".	The system successfully redirects to Google Maps according to the intended location point.	Valid

Table 5
Results of k-means clustering.

District	Number of Crime Cases (Per Year)			
	2020	2021	2022	Cluster
Pabean Cantian	73	83	57	1
Mulyorejo	71	65	53	1
Krembangan	72	66	59	1
Pakal	60	66	68	1
Sambikerep	56	91	49	1
Simokerto	74	97	52	1
Dukuh Pakis	57	81	51	1
Jambangan	72	82	38	1
Tambaksari	82	85	47	1
Wiyung	77	78	28	1
Bubutan	76	72	56	1
Benowo	58	72	39	1
Kenjeran	66	57	41	2
Bulak	62	62	50	2
Tenggilis Mejoyo	75	55	32	2
Sukolilo	78	60	28	2
Gubeng	59	43	47	2
Gunung Anyar	58	62	32	2
Semampir	72	57	50	2
Sawahan	77	57	35	2
Karang Pilang	72	68	38	2
Gayungan	70	56	52	2
Tandes	64	66	48	2
Lakarsantri	75	47	50	2
Wonocolo	65	52	40	2
Genteng	88	56	48	3
Asemrowo	85	68	39	3
Rungkut	96	92	50	3
Wonokromo	90	77	55	3

(*continued on next page*)

Table 5. (*continued*)

District	Number of Crime Cases (Per Year)			
	2020	2021	2022	*Cluster*
Suko Manunggal	86	72	44	3
Tegalsari	98	62	54	3

Table 6.
Number of regional clusters in Surabaya.

Cluster	Total	Category
1	12	High crime rate
3	6	Medium crime rate
2	13	Low crime rate

5. CRediT Authorship Contribution Statement

Supangat: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing - Original Draft, Writing - Review & Editing and Funding acquisition. **M. Mudhafiq Sholiq:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing - Original Draft, Writing - Review & Editing and Funding acquisition.

6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

7. Acknowledgments

The authors would like to thank the anonymous referees for their helpful comments and suggestions.

8. Data Availability

Data will be made available on request.

9. Funding

No funds, grants, or other support were received.

10. Ethical Approval

Ethical approval No patient-identifying parts in this paper were used or known to the authors. Therefore, no ethical approval was requested.

11. References

Anshori, M. H., & Misbachudin, M. (2017). Rancang Bangun Simulasi Safety Starting System pada Mobil L300. *Polhasains: Jurnal Sains Dan Terapan Politeknik Hasnur*, *5*(2), 27–31. https://ejournal.polihasnur.ac.id/index.php/phssains/article/view/208

Arfani, F., & Nashrullah, H. (2022, March). Polrestabes Surabaya ungkap 58 kasus kejahatan jalanan. *Antara News*. https://www.antaranews.com/berita/2741061/polrestabes-surabaya-ungkap-58-kasus-kejahatan-jalanan

Arsista, G. (2022). *Alert! Kasus Curanmor di Surabaya Meningkat*. Suara Surabaya.

Astuti, Y. (2018). Program Smart City Kota Makassar (Studi Pemanfaatan Closed Circuit Television di Kota Makassar). In *Universitas Muhammadiyah Makassar*. https://digilibadmin.unismuh.ac.id/upload/4178-Full_Text.pdf

Bindosano, K. A., Yudono, A., & Hasyim, A. W. (2022). Arahan Pengurangan Kriminalitas di Kota Jayapura. *Planning for Urban Region and Environment Journal (PURE)*, *11*(4), 231–240.

Gunawan, M., & Aditya, T. (2019). Analitik Geovisual Pola Kejadian dan Pergerakan Kriminalitas Berbasis Data Media Sosial. *Prosiding Seminar Nasional GEOTIK*, 335–345.

Hapsari, D. P. T., & Widodo, E. (2017). Pengelompokan Daerah Rawan Kriminalitas di Indonesia Menggunakan Analisis k-Means Clustering. *Prosiding SI MaNIs (Seminar Nasional Integrasi Matematika Dan Nilai Islami)*, *1*(1), 147–153. http://conferences.uin-malang.ac.id/index.php/SIMANIS/article/view/52

Himawan, A. (2023). Optimalisasi Pemolisian Masyarakat Guna Memberikan Rasa Aman dan Rasa Nyaman pada Masyarakat, Sudut Pandang Ilmu Psikologi. *Jurnal Litbang Polri*, *26*(1), 30–40. https://jlp.puslitbang.polri.go.id/jlp/LitbangPOLRI/article/view/197

Kendall, K. E., & Kendall, J. E. (2011). *Systems Analysis and Design* (8th ed.). Prentice Hall.

Larose, D. T., & Larose, C. D. (2014). Discovering Knowledge in Data: An Introduction to Data Mining. In *John Wiley & Sons*. https://doi.org/10.1002/9781118874059

Mulyani, H., Nurjaman, J., & Nugraha, M. (2020). Geographic Information Systems for Crime Prone Areas Clustering. *JOIN (Jurnal Online Informatika)*, *5*(2), 227–232. https://doi.org/10.15575/join.v5i2.599

Nurjoko, N., Dwirohayati, D., & Sudibyo, N. H. (2020). Sistem Informasi Pemetaan Wilayah Rawan Kriminalitas Polresta Bandar Lampung Menggunakan K-Means Clustering. *Teknika*, *14*(2), 127–135.

Nurman, R. M. (2007). istem Informasi Pemetaan Profil Kriminalitas Berbasis Web (Studi Kasus: Kejahatan Konvensional Kota Bogor). In *Institut Pertanian Bogor*.

Prasetyo, E. (2014). *Data Mining : mengolah data menjadi informasi menggunakan Matlab*. Andi Publisher.

Rahayu, S., Nugrahadi, D. T., & Indriani, F. (2014). Clustering Penentuan Potensi Kejahatan Daerah di Kota Banjarbaru dengan Metode k-Means. *KLIK (Kumpulan JurnaL Ilmu Komputer)*, *01*(01), 33–45.

Rijali, A. (2018). Analisis Data Kualitatif. *Alhadharah: Jurnal Ilmu Dakwah*, *17*(33), 81–95. https://doi.org/10.18592/alhadharah.v17i33.2374

Setiawan, I., Dede, M., Sugandi, D., & Widiawaty, M. A. (2019). Investigating Urban Crime Pattern and Accessibility Using Geographic Information System in Bandung City. *KnE Social Sciences*, 535–548. https://doi.org/10.18502/kss.v3i21.4993